

UNIVERSITÉ D'ALGER
FACULTÉ MIXTE DE MÉDECINE ET DE PHARMACIE D'ALGER
ANNÉE 1938 — N° 13

Contribution à l'Etude
des
STATISTIQUES MÉDICALES

Les Erreurs de Base



THÈSE

pour le Doctorat en Médecine
présentée et soutenue publiquement le 3 Juin 1938
par

Charlotte RICARD

NÉE LE 25 NOVEMBRE 1910 A ORAN (DÉP. D'ORAN)
ANCIENNE EXTERNE DES HOPITAUX
CERTIFICAT DE MICROBIOLOGIE



JURY

MM. SENEVET, <i>Professeur de parasitologie et Zoologie médicale</i>	Président
PINOY, <i>Professeur de bactériologie</i>	Juges
PORTES, <i>Professeur de physique médicale</i>	
BENHAMOU, <i>Agrégé (Médecine)</i>	

Contribution à l'Etude
des
STATISTIQUES MÉDICALES

Les Erreurs de Base



THÈSE

pour le Doctorat en Médecine

présentée et soutenue publiquement le 3 Juin 1938

par

Charlette RICARD

NÉE LE 25 NOVEMBRE 1910 A ORAN (DÉP. D'ORAN)

ANCIENNE EXTERNE DES HOPITAUX

CERTIFICAT DE MICROBIOLOGIE



JURY

MM. SENEVET, <i>Professeur de parasitologie et Zoologie médicale</i>	Président
PINOY, <i>Professeur de bactériologie</i>	Juges
PORTES, <i>Professeur de physique médicale</i>	
BENHAMOU, <i>Agrégé (Médecine)</i>	

UNIVERSITÉ D'ALGER

FACULTÉ MIXTE DE MÉDECINE ET DE PHARMACIE D'ALGER

DOYEN	MM. LEBLANC (F. F. I. U)
DOYEN HONORAIRE	HERAIL (F. I. U)
ASSESEUR	COSTANTINI (* F. I. U)

PROFESSEURS

Clinique médicale	AUBRY (F. F. I. U)
Clinique gynécologique	CABANES (F. F. I. U)
Clinique chirurgicale	COSTANTINI (* F. I. U)
Botanique et matière médicale	FOURMENT (I. U. O. O. U)
Clinique médicale et hygiène infantile	GILLOT (I. U)
Médecine légale	GIRAUD (O. F. F. I. U)
Clinique obstétricale et puériculture du premier âge	LAFFONT (O. F. F. I. U)
Anatomie	LEBLANC (F. F. I. U)
Thérapeutique	LEBON (F. F. I. U)
Clinique chirurgicale infantile et d'orthopédie	LOMBARD (* F. I. U)
Chimie biologique et médicale	GIBERTON (I. U)
Pharmacie	MUSSO (* I. U)
Bactériologie	PINOY (* I. U)
Clinique psychiatrique	POROT (I. U)
Physique médicale	PORTES (F. F. I. U)
Anatomie pathologique	MONTPELLIER I. U)
Clinique des maladies des pays chauds, des maladies syphilitiques et cutanées	RAYNAUD (F. F. I. U)
Parasitologie et zoologie médicale	SENEVET (F. I. U)
Physiologie	TOURNADE (O. F. F. I. U)
Histologie et Embryologie	COURRIER (* F. I. U)
Chimie générale, pharmaceutique et toxicologique	WUNSCHENDORFF (I. U)
Clinique ophtalmologique	TOULANT (* F. U)
Hygiène, hydrologie et climatologie	N...

PROFESSEURS HONORAIRES

ARGAUD (F. F. I. U)	HERAIL (F. I. U)
POUJOL (F. I. U)	WEBER (* F. I. U)

CHARGES DE COURS

Clinique d'oto-rhino-laryngologie	MM. ABOULKER (F. F. I. U)
---	---------------------------

AGREGES

Médecine	BENHAMOU (* F. A. U)
Histologie	LACROIX (A. U)
Chirurgie	N...
Obstétrique	SABADINI
Anatomie pathologique	MOGGI (* I. U)
Anatomie	HOUEL (* F. I. U)
Chimie médicale	N...
Histoire naturelle pharmaceutique	RIBET (F. A. U)
Pharmacie	N...
	ROQUES (I. U)
	MONNET'

AGREGES LIBRES

Physiologie	BATTEZ
Chirurgie	FERRARI (I. U)
	DUBOUCHER (I. U)

NOTA. — La Faculté n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

SERMENT D'HIPPOCRATE

En présence des Maîtres de cette Faculté, de mes chers condisciples et selon la tradition d'Hippocrate, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine. Je donnerai mes soins gratuits à l'indigent, et n'exigerai jamais un salaire au-dessus de mon travail. Admis dans l'intérieur des maisons, mes yeux ne verront pas ce qui s'y passe, ma langue taira les secrets qui me seront confiés et mon état ne servira pas à corrompre les mœurs ni à favoriser le crime.

Respectueux et reconnaissant envers mes Maîtres, je rendrai à leurs enfants l'Instruction que j'ai reçue de leurs pères.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.

STAGES

Année 1932 :

Clinique Chirurgicale	Professeur COSTANTINI
Clinique médicale	Professeur AUBRY

EXTERNAT

1. Semestre 1933 :

Clinique ophtalmologique ...	Professeur CANGE
------------------------------	------------------

2. Semestre 1933 :

Clinique chirurgicale	Professeur COSTANTINI
-----------------------------	-----------------------

1. Semestre 1934 :

Clinique gynécologique	Professeur CABANES
------------------------------	--------------------

2. Semestre 1934 :

Clinique dermatologique et maladies des pays chauds ..	Professeur RAYNAUD
---	--------------------

2. Semestre 1935 :

Clinique médicale	Professeur AUBRY
-------------------------	------------------

1. Semestre 1936 :

Clinique de médecine infantile	Professeur GILLOT
--------------------------------	-------------------

2. Semestre 1936 :

Clinique ophtalmologique	Professeur TOULANT
-------------------------------	--------------------

1. Semestre 1937 :

Consultation de médecine in- fantile	Professeur GILLOT.
---	--------------------

STAGES DE SPECIALITES

Clinique obstétricale	Professeur LAFFONT
Clinique de neuro-psychiâtrie	Professeur POROT
Maladies contagieuses	Docteur LEMAIRE
Oto-rhino-laryngologie	Docteur ABOULKER

Je dédie cette thèse :

A CEUX DES MIENS QUI NE SONT PLUS

A MON MARI, LE DOCTEUR EMILE RICARD

*En gage de mon immense
amour.*

A MON PERE ET A MA MERE

*En témoignage de profonde
affection.*

A MA SŒUR, MARCELINE ANTOINE

Etudiante en Médecine

A MES ONCLES ET A MES TANTES

A MA COUSINE, LE DOCTEUR RENÉE ANTOINE
Ancien chef de Clinique à la Faculté

A MES BEAUX-PARENTS

A MADAME ET AU DOCTEUR GEORGES RICARD

A MON PRESIDENT DE THESE

MONSIEUR LE PROFESSEUR SENEVET

Nous le remercions de l'honneur qu'il nous fait en présidant aujourd'hui notre jury, après avoir inspiré ce travail.

A MES JUGES

MONSIEUR LE PROFESSEUR PINOY

MONSIEUR LE PROFESSEUR PORTES

MONSIEUR LE PROFESSEUR AGRÉGÉ BENHAMOU

Qui furent aussi nos maîtres.

A MES MAITRES DE LA FACULTE

A MES MAITRES DE L'HOPITAL

A MONSIEUR LE PROFESSEUR AGRÉGÉ BENHAMOU

Nous le remercions de la sympathie qu'il nous a témoignée en fin de nos études et de l'accueil qu'il nous a fait dans son service.

Nous le prions de vouloir bien trouver ici l'expression de notre vive gratitude.

A MONSIEUR LE PROFESSEUR GILLOT

Nous garderons toujours le souvenir de la bienveillance avec laquelle il nous a reçue, lors de nos stages d'externat, dans son service.

Nous le prions d'agréer l'assurance de toute notre reconnaissance.

A MES AMIS ET CAMARADES D'ETUDE :

MADAME HÉLÈNE LAMBERT

MADemoisELLE LISE Ayme

MADAME ANDRÉE PIRSON

MADemoisELLE ANNE-MARIE DE BONNECAZE

MADemoisELLE ALICE DUVALLET

DOCTEUR ANDRÉ MONPERE

DOCTEUR HÉLÈNE CORIZZI

DOCTEUR MARIE-JEANNE CORIZZI

DOCTEUR CLAUDE SOLACROUP

DOCTEUR MAURICETTE PUJOL

DOCTEUR GEORGES AMARANTINI

DOCTEUR SUZY BEFFEYTE

DOCTEUR MADELEINE ROIG

DOCTEUR HUGUETTE XIMENES

DOCTEUR PIERRE FRIMIGACCI

DOCTEUR PIERRE MINICONI

MONSIEUR EDMOND VIGUIE

Avocat à la Cour d'Appel

Avant-Propos

La médecine actuelle tend de plus en plus vers la traduction mathématique, en formules précises, de faits qu'il y a seulement cinquante ans l'intuition, le flair clinique et l'expérience étaient seuls à se disputer.

De nos jours on veut être précis. Le moment n'est plus où la main du praticien remplaçait la thermométrie et où la sensation de chaleur tenait lieu de degré centigrade. L'âge du chiffre est arrivé, et, tous les jours, dans chaque département de la science, on tend à prouver la valeur de telle ou telle expérience par l'interprétation de données numériques basées sur des faits observés et non plus sur des sentiments profonds, plus intuitifs qu'objectifs. L'époque des approximations est révolue.

Les praticiens qui désavouent le plus cette façon de « mathématiser » la médecine sont cependant d'accord sur le fait qu'une présentation en chiffres est indispensable pour objectiver les succès et les échecs de telle ou telle expérience, et surtout pour établir des comparaisons entre différentes expériences ou entre les mêmes expériences à des époques différentes.

Le procédé employé est la statistique médicale :

C'est une science qui a pour objet de fournir des documents exacts sur les diverses parties de l'activité médicale en recueillant, classant, comptant tous les

faits de même ordre qui se produisent en un temps donné.

Ces statistiques emploient les mathématiques pour arriver aux résultats cherchés; c'est là, en même temps qu'un de leurs avantages, un inconvénient pour ceux dont l'esprit n'est pas particulièrement orienté vers la science des chiffres.

Le Docteur Bradford Hill¹, dans son introduction à la série d'articles qu'il fit paraître dans « The Lancet », chaque semaine, pendant les mois de Janvier, Février, Mars et Avril 1937, s'exprime en ces termes en parlant de statistiques médicales :

« Les statistiques sont de curieuses choses, elles sont un exemple dans lequel l'usage ou l'abus des méthodes mathématiques tend à introduire une véritable émotion dans les esprits non mathématiques. La cause en est que les statisticiens appliquent au problème auquel nous nous intéressons, une technique que nous ne comprenons point. »

Pour permettre au praticien de mieux comprendre la technique des statistiques, nous avons essayé, en relevant quelques causes d'erreur, de lui montrer combien les statistiques étaient souvent erronées, tant à cause de leur construction que de leur interprétation.

Trop souvent en effet, comme nous le dit Auguste LUMIERE — « Les conditions dans lesquelles les statistiques sont établies, ont sur les résultats qu'elles enregistrent une influence si considérable que les conclusions auxquelles elles aboutissent se trouvent parfois en opposition flagrante avec les faits les plus patents ; aussi a-t-on maintes fois prétendu, fort justement d'ailleurs, que l'on peut prouver par la démo-

graphie tout ce que l'on désire démontrer et voire même à volonté le contraire. »

Nous avons pris des exemples dans les différentes publications médicales pour illustrer à la fois les différents problèmes statistiques que le praticien aura à résoudre et aussi les causes d'erreur qu'il lui faudra éviter, pour bâtir sa statistique et surtout pour l'interpréter.

Nous avons été obligés au cours de ce travail de citer et même de reproduire certaines de ces publications. Nous prions leurs auteurs de croire que nous n'avons voulu faire aucune personnalité et que c'est le hasard seul qui a présidé à l'analyse de leurs articles plutôt que d'autres.

C'est un travail ingrat que de réfuter les conclusions d'auteurs dont l'autorité en dehors de l'application des méthodes statistiques est par ailleurs incontestée.

Nous citerons ces mots de Bradford Hill qui traduisent bien notre pensée :

« Aucun statisticien, malgré des opinions opposées, n'aime réfuter les conclusions d'autres travailleurs, surtout quand il reconnaît qu'une grande somme de patience et de travail a été dépensée dans la récolte des faits. »

Nous avons essayé, au cours de cette étude des erreurs statistiques, de restreindre au minimum la partie mathématique pure, à l'exception de quelques formules abstraites que nous n'avons pu éviter pour la démonstration du bien fondé de nos critiques.

Les textes anglais que nous avons traduits sont reproduits dans ce travail. Nous avons pensé que la compétence de Bradford Hill était tellement au-dessus

de la nôtre, que rien ne pouvait remplacer l'expression de sa pensée dans l'étude de certains problèmes.

Enfin, et pour résumer, nous nous proposons dans ce travail de réhabiliter la statistique médicale correctement établie et de donner un démenti à Disraëli qui a dit quelque part :

— « Il existe trois formes de mensonges :

Le mensonge direct ;

La prétérition ;

Et la statistique. »

Historique

C'est Achenvall, professeur de droit public à l'Université de Gœttingue, qui a donné à la statistique son nom vers le milieu du 18^e siècle. Il la définissait :

« La connaissance approfondie de la situation (Status) respective et comparative de chaque état. »

Schloeser, qui lui succéda à l'Université, écrivait que : « la statistique a pour but de faire connaître tous les objets dont se compose la puissance d'un Etat. »

Il ajoutait : « L'Histoire est la statistique en mouvement et la statistique est l'Histoire au repos. »

Moreau dans ses « Eléments statistiques » dit :

« La statistique est la science des faits sociaux exprimés par des termes numériques. »

Ross et W. Stott publient en 1911, le 24 octobre, dans « The Annals of Tropical Medicine and Parasitology » leurs tables « of statistical errors ».

Bradford Hill D. Sc. PH; D. publie dans « the Lancet », les « Principes of medical statistics. »

Mendel dans ses lois de Génétique, peut être considéré comme ayant participé par ses calculs de probabilité aux progrès des sciences statistiques.

DEFINITION

D'étymologie grecque :

Statizein qui signifie établir, constater, c'est un mot

qui a été rattaché par Eichhoff à la racine sanscrite Sthas qui signifie : placer, fixer.

La statistique se définit :

Science qui a pour objet le groupement méthodique des faits sociaux qui se prêtent à une évaluation numérique.

Ou :

Science qui a pour objet de fournir des documents exacts sur les diverses parties de la vie d'un peuple, en : — recueillant,

— classant,

— comptant tous les faits de même ordre, (naissances, décès, mariages, maladies...), qui se produisent dans un temps donné.

Ou encore :

« Science qui déduit la situation générale du dénombrement de tous les résultats, de tous les objets similaires qui peuvent être utiles aux membres de la société. »

(*Charles Dupin.*)

On peut aussi donner le nom de statistique au dénombrement leucocytaire ou hématique.

En effet, la formule leucocytaire évalue, en partant d'un échantillon du sang total, le pourcentage d'éléments blancs de toute la masse sanguine. Il en est de même pour les numérations globulaires ou le pourcentage des parasites du sang d'un sujet infesté.

METHODES DE LA STATISTIQUE

Il y a deux manières d'obtenir l'expression numérique des faits sociaux :

— l'une : Méthode naturelle ou d'exposition ;

— l'autre : obtenir les résultats à l'aide des ressources qu'offre le calcul.

Première méthode :

Méthode naturelle ou d'exposition.

Elle consiste à recueillir un à un tous les faits numériques qui constituent les éléments de l'ordre de faits que l'on se propose de connaître : à les grouper ensemble, à les coordonner et au besoin à les réduire pour en avoir la mesure sous une formule plus comode, le tout sans les altérer.

Deuxième méthode :

Calcul des probabilités.

Il se propose de déterminer le nombre de chances qui peuvent amener une éventualité donnée, ou en d'autres termes, de formuler les chances calculables des événements humains.

PRESENTATION DES STATISTIQUES

Quand les résultats sont obtenus, les statisticiens les présentent sous la forme de tableaux ou de diagrammes ; le plus souvent ces diagrammes accompagnent les tableaux ; les résultats sont livrés ainsi à l'interprétation du public, sans autre explication que la méthode d'obtention des résultats et la taille des échelles adoptées par les auteurs.

PLAN

Nous avons divisé l'étude des erreurs que l'on rencontre le plus fréquemment dans l'établissement des statistiques, en trois parties. La première traitera des erreurs dues à des causes volontaires ; la seconde analysera les erreurs produites par une discrimination

insuffisante ; la troisième visera spécialement les erreurs mathématiques, que l'on pourra grouper en deux parties bien différentes :

1° Erreurs mathématiques par l'application irréflechie de l'arithmétique au calcul des index totaux.

2° Erreurs mathématiques par l'application irréflechie de l'arithmétique au calcul des moyennes.

Des observations tirées de la littérature médicale française ; d'autres empruntées à des auteurs étrangers, illustreront les différentes erreurs que le praticien non averti commet encore trop fréquemment.

CHAPITRE I

Les Causes d'Erreur
Volontaires

A. — CAUSES D'ERREUR PAR L'APPLICATION DU SECRET PROFESSIONNEL.

L'emploi de la démographie est à la base de toute médecine sociale. De cette démographie dépendront les conclusions que le praticien tirera de ses résultats. Il est donc indispensable que ces résultats ne soient pas erronés et qu'ils aient été « enregistrés d'une façon précise, réunis et comparés en tenant compte de tous les facteurs qui les conditionnent. »

Si la démographie ne peut p'lus assumer cette fonction, le médecin sociologue ne peut agir qu'au hasard; son rôle est condamné à l'empirisme, et sa mission devient illusoire.

Les statistiques de mortalité pour les différentes maladies sont nécessairement à la base de toute action prophylactique efficace. Nous avons noté que la documentation diagnostique qui préside à sa construction correcte, devient de jour en jour p'lus précaire, pour ne pas dire de plus en plus insuffisante.

La cause en est : le respect abusif du secret professionnel.

Loin de nous la pensée de vouloir aller à l'encontre de cette respectable tradition, l'une des plus fières prérogatives de l'exercice de la médecine, mais dont l'application, et surtout l'interprétation outrancière, nuit à l'intérêt général.

Il est certain que si tous les médecins s'abstenaient de mentionner le diagnostic sur leurs certificats de décès, il ne serait plus possible de savoir « de quoi on meurt en France. »

Heureusement que beaucoup d'entre eux se ren-

dent compte de l'importance des renseignements qu'ils fournissent et que les certificats de décès ne sont pas tous affublés de la mention :

« Décès pour cause indéterminée. »

De plus, et quand on y réfléchit davantage, l'abstention des causes de décès ne joue pas également pour toutes les causes de la léthalité. Le public considère à tort en effet, la tuberculose, la syphilis, le cancer, la vésanie, l'alcoolisme etc... comme des façons moins avouables que d'autres de mourir ; et les médecins seront évidemment peu enclins à vouloir mécontenter leurs clients sur ces points, en affirmant par ceriticat, la véritable cause de la mort.

Voilà donc la tuberculose, la syphilis, le cancer, la folie et l'alcoolisme, qui, sur les statistiques de mortalité, sont en voie de disparition.

Le procédé et d'autant plus dangereux que, très souvent, le diagnostic de tuberculose par exemple, est remplacé par le diagnostic d'autres affections, et « leurs certificats porteront volontiers la mention de pneumonie, d'affections cardiaque, de fièvre typhoïde », ce qui, en augmentant les autres causes de décès, réduit encore le pourcentage de la mortalité tuberculeuse.

Sans vouloir médire des comités nationaux de défense contre la tuberculose ou la syphilis, nous voudrions simplement faire remarquer que si on note une régression notable des cas de tuberculose ou de syphilis dans les statistiques françaises, deux causes y ont collaboré :

1° La lutte anti-tuberculeuse ou anti-syphilitique et
2° l'abstention du diagnostic dans les ceritcats pour tuberculose ou pour syphilis.

Les renseignements fournis par le docteur Vigne, Directeur du bureau d'hygiène de la ville de Lyon, au sujet des certificats sans diagnostic pendant les années 1923, 1933 et le premier semestre 1934, nous donne :

années	Décès totaux	Décès sans indication	Décès par tuberculos.
1923	8.073	635 soit 7,8 %	1.032
1933	7.918	2.342 soit 29,57%	457
1934	4.172	1.621 soit 38,85%	245

D'après ce tableau on a le droit de conclure que très bientôt la mortalité par tuberculose sera extrêmement rare, mais non que la cause de cette régression du fléau est due aux seules méthodes prophylactiques instaurées.

Par contre, dans d'autres villes où le spectre du secret professionnel n'a pas détourné les médecins de leurs devoirs envers la nation, le taux de mortalité par tuberculose est resté sensiblement le même.

Que penser de ces contradictions, sinon que les statistiques faussées à la base par une documentation diagnostique insuffisante ne signifient plus rien ?

Quel est remède ?

Concilier le point d'honneur qu'a le médecin à user d'une prérogative que lui a donné le serment d'Hippocrate et l'intérêt même de l'Etat.

Cette formule est plus facile à écrire qu'à mettre en pratique ; peut-être que la suppression de tous les in-

termédiaires administratifs, non liés par le secret professionnel, et qui jalonnent la voie du certificat de décès jusqu'au bureau central de la statistique, inclinerait sentimentalement les médecins, à confier plus volontiers le vrai diagnostic de l'affection dont sont morts leurs patients.

Mais ceci est une anticipation...

B. — LES PUBLICATIONS PREMATUREES

*« Le plus grand dérèglement
de l'esprit est de croire aux choses
parce qu'on veut qu'elles
soient. »*

BOSSUET.

(Epitaphe empruntée par Pasteur pour l'inscrire sur son livre mémorable « Etudes sur la Bière. »)

Au sujet des publications prématurées, nous abandonnerons l'esprit scientifique de ce travail, dans le choix des publications que l'on trouve cependant en assez grand nombre dans la littérature médicale.

Pour ne blesser aucune susceptibilité, nous choisirons nos exemples dans les œuvres romancées de Georges Duhamel et de Sinclair Lewis.

Le professeur Rhoner dans « *Les Maîtres* », de l'auteur de la chronique des Pasquiers, fait passer la probité scientifique après la confirmation de ses idées personnelles :

Pasquier dit, en parlant du Professeur Rhoner :

— « Et soudain, j'ai compris qu'il était tout entier en

proie à l'idée fixe, qu'il ne cherchait pas la vérité, mais seulement la confirmation de ses songeries et qu'il fallait faire en sorte de trouver cette confirmation, coûte que coûte; qu'il allât interroger les tissus, de telle manière que les tissus tourmentés répondraient n'importe quoi....

Pages 229, 300 et 301 :

— « Monsieur Nicolas Rhoner a publié ces jours derniers un mémoire sur le streptococcus Rohneri et sur la maladie nouvelle dite maladie de Rhoner. Il paraît que Catherine Houdoire est morte d'une endocardite et d'une néphrite caractéristiques toutes deux et vérifiées au microscope sur les coupes histologiques.

Les Maîtres, de Georges Duhamel, de l'Académie de médecine et de l'Académie Française.

Arrowsmith nous offre des exemples de publications sur commande et coûte que coûte, dans ces lignes :

Angus à Arrowsmith :

« Martin dit-il, je suis heureux de te voir rester en contact avec la science, mais à ta place, j'évitais de dépenser tant d'énergie à satisfaire une simple curiosité.

Le docteur Rouncefield m'en parlait l'autre jour. Nous aurions tous deux plaisir à te laisser faire toutes les recherches qu'il te plairait, mais nous voudrions que ces recherches tendent à quelques résultats pratiques. Si tu pouvais par exemple établir une table de numération des globules sanguins dans quelques centaines de cas d'appendicite et la publier quelque

part, ce travail en soi mènerait à quelque chose, et, si tu t'arrangeais en même temps pour y faire mention de la Clinique, nous en recevions aussi quelque crédit et peut-être nous serait-il possible de porter à trois mille dollars tes appointements annuels... (Page 372 et 373).

Et encore :

« J'ai pensé, dit-il, Arrowsmith. (Il avait en effet l'air d'un homme qui avait beaucoup pensé) La découverte de d'Herelle n'a pas suscité dans les masses l'intérêt que j'eusse cru qu'elle provoquerait. Si ce savant s'était trouvé des nôtres, j'aurais pourvu à ce qu'il obtint l'attention qu'il mérite. Mais peut-être pouvons-nous faire encore aujourd'hui quelque chose de sa découverte. Vous avez à ce que je comprends, poussé maintenant assez loin ce que Gottlieb appelle « La recherche fondamentale ». Il est temps pour vous d'appliquer le phage à la guérison des maladies. Instituez des expériences en vue de traiter par ce moyen la pneumonie, peut-être aussi la peste, la typhoïde, et, dès que ces expériences seront en train, faites des essais pratiques dans les hôpitaux. En voilà assez de cet étalage de vanité, et de ce gaspillage de temps. Guérissons enfin quelqu'un.

Martin n'était pas tout fait exempt de la crainte qu'il ne fut renvoyé de l'Institut s'il désobéissait au Patron....»

Et plus loin :

A toutes les objections de Martin, Pickerbaugh répondait :

« Et qu'importe que mes statistiques ne soient pas toujours exactes, que ma publicité et les divertisse-

ments que j'offre au peuple semblent vulgaires à certains ? Tout cela est finalement bienfaisant, et par suite bon en soi. Quel'es que soient nos méthodes, si nous obtenons par elles què les gens respirent plus d'air pur, qu'ils tiennent un peu plus propres les cours de leurs maisons et s'adonnent un peu moins à l'alcool, nous sommes de ce fait amplement justifiés. » (page 303.)

Enfin, Arrowsmith lui-même se prend à douter de ses propres sentiments et il dit :

« Au fait, ce'a a-t-il après tout tant d'importance ? La vérité, cette vérité froide, hostile... importe-t-elle si grandement ?... Mais se souci de précision scientifique ne serait-il pas chez moi une marotte, comme l'est par exemple chez d'autres, la passion du golf ?... » (pages 303-304.)

Arrowsmith, de Sinclair Lewis
Préface de Jacques Risler, chef
du laboratoire de physiologie
Charles Henry à la Sorbonne.

Un autre exemple, celui-là tiré d'un journal médical
« Traité des maladies parasitaires de Neumann (au chapitre de la trichinose) page 691 :

« Un fait ressort des chiffres fournis par les américains : c'est qu'ils accusent beaucoup moins de trichine depuis que les populations de l'Europe se sont émues de leurs envois infestés. »

N'ayant pu contrôler le bien-fondé de cette remarque, plutôt désobligeante pour les statisticiens américains, nous laissons à l'auteur la responsabilité d'un tel jugement.

Nous n'insisterons pas davantage sur les causes d'erreur qu'une telle façon de comprendre la vérité scientifique entraîne, dans l'interprétation des résultats.

CHAPITRE II

Causes d'Erreur par Discrimination insuffisante

Parmi les causes d'erreur les plus fréquentes, ce sont celles qui proviennent d'une discrimination insuffisante, qui doivent tout d'abord retenir notre attention.

Qu'est-ce que la « discrimination insuffisante ? ».

Du latin « discrimen » qui signifie caractère distinctif, la discrimination est l'acte psychologique qui permet de distinguer, de discerner dans plusieurs objets les caractères communs ou différents.

La « discrimination insuffisante » est donc un défaut de discernement entre différents caractères.

Quand deux objets ont un caractère commun, la discrimination les rapproche, les isole des objets qui les entourent, en fait une classe à part, une catégorie spéciale.

Mais, si par ailleurs les deux objets à caractères communs sont en certains points très différents, la discrimination doit en tenir compte dans le groupement de ces objets en une espèce particulière.

C'est ainsi que très souvent, les praticiens qui établissent des statistiques sont attirés par les caractères communs des sujets observés, mais ne tiennent aucun compte des caractères distinctifs propres à chacun d'eux.

Par exemple : Tous les sujets peuvent présenter une cuti-réaction positive (caractère commun), mais les uns sont des enfants, les autres des nourrissons, d'autres enfin des adultes citadins ou campagnards, autant de caractères distinctifs que la discrimination insuffisante n'a pas retenus, mais qui ont cependant

leur importance dans l'interprétation des pourcentages ou de l'index tuberculinique total établi.

Nous nous sommes inspirés du travail de Bradford Hill dans l'étude des différentes sortes de discriminations insuffisantes.

Nous envisagerons successivement :

- 1° — Le Mélange de chiffres non comparables ;
- 2° — La négligence de la période d'exposition au risque ;
- 3° — L'absence de l'exposé au risque ou standard de comparaison ;
- 4° — L'association directe ou indirecte ;
- 5° — La notion de choix des échantillons.

1. — MELANGE DE CHIFFRES NON COMPARABLES

Ici la discrimination n'a pas su discerner que la différence de sexe pouvait fausser les résultats de l'application d'un traitement à des hommes et des femmes atteints d'une même maladie, mais à pourcentage de mortalité différent.

Exemple tiré de Bradford Hill :

Supposons que dans une maladie particulière, le taux de la mortalité chez les femmes soit plus élevé que chez les hommes :

- 40 % chez les premières ;
- 20 % chez les seconds.

Une nouvelle forme de traitement est appliquée à 80 hommes et à 40 femmes.

L'observation porte sur 30 hommes et 60 femmes. Le nombre des décès observés sur les 120 individus

auxquels on a donné le nouveau traitement est de 32, donnant un pourcentage de 26,7 %, tandis que le chiffre des décès observés parmi les 90 individus sur lesquels a porté spécialement le contrôle est de 30, soit un pourcentage de 33,3 %.

A première vue cette comparaison suggère que le nouveau traitement a quelque valeur. En fait, cette conclusion est injustifiée parce qu'on n'a pas comparé des éléments comparables : « Like with like ».

Le nombre total des décès parmi les 120 malades doit être influencé par la proportion des deux sexes présents dans chaque lot. Hommes et femmes en fait ne sont pas également représentés dans le lot traité et dans celui que l'on a pris comme contrôle.

Le tableau ci-dessous montre clairement la faute :

	Hommes	Femmes	les deux
% des décès normaux	20 %	40 %	
Malades traités	80	40	120
Décès dans le groupe traité	16	16	32
% décès groupe traité	20 %	40 %	26,7 %
Nombre de malades contrôlés	30	60	90
Décès dans ce groupe	6	24	30
% Décès dans ce groupe ..	20 %	40 %	33 %

La comparaison « Like with like », c'est-à-dire avec hommes avec hommes et femmes avec femmes, montre que le traitement n'a pas de valeur puisque les pourcentages de décès des groupes hommes et femmes, traités et non traités, sont identiques et égaux au taux normaux des décès chez les hommes et chez les femmes.

Il faut donc, avant d'établir une comparaison entre deux résultats statistiques, être certain que les deux résultats comparés ont été établis à partir d'éléments qu'une discrimination suffisante a jugé comparables.

C'est ainsi qu'on ne doit pas comparer les index tuberculiniques des enfants et des adultes ; et qu'il ne faut sous aucun prétexte se baser, dans l'évaluation du pourcentage de cuti-réactions positives, sur un index établi en additionnant (ce qui revient à vouloir rendre comparables), des chiffres qui mesurent le degré de sensibilisation à la tuberculine des enfants et des adultes.

Quand on a su discerner les éléments qui peuvent être comparés, il reste encore d'autres causes d'erreur par discrimination insuffisante. Deux lots de sujets peuvent présenter tous les caractères communs nécessaires à un rapprochement statistique et différer seulement par un point : le temps d'exposition au risque. Cette erreur est commise surtout dans l'étude de l'efficacité des prophylaxies ; un exemple tiré de Bradford Hill nous montre combien l'interprétation des résultats peut être variable avec ou sans cette notion.

2. — NEGLIGENCE DE LA PERIODE D'EXPOSITION AU RISQUE

La discrimination est également insuffisante quand

elle néglige de discerner parmi tous les sujets observés le temps d'exposition au risque de contamination.

Une autre cause d'erreur ressort de la comparaison de sujets vaccinés et non vaccinés, en négligeant le temps pendant lequel les individus sont exposés, d'abord dans un groupe (non vaccinés par exemple) puis dans l'autre (vaccinés).

Supposons que le 1^{er} janvier 1936, il y ait dans une région trois cents sujets vaccinés et mille non vaccinés.

Le nombre des atteintes morbides est compté pendant un an, c'est-à-dire jusqu'au 1^{er} janvier 1937.

Ce nombre est comparé avec les atteintes annuelles des années précédentes. Ceci est exact.

Mais, si, comme cela arrive souvent en pratique, des gens sont vaccinés en cours d'année d'observation, la comparaison ne peut s'établir, car les uns, étant vaccinés en Février, en Avril ou en Mai, ne seront pas restés un an non vaccinés ou un an vaccinés, mais : un an moins un mois pour ceux de Février, un an moins 3 mois pour ceux d'Avril etc...

Supposons que le 1^{er} janvier 1936, nous observions cinq mille personnes non vaccinées.

Admettons que trois cents d'entre elles soient vaccinées au 1^{er} Avril, six cents au 1^{er} Juillet, cent au 1^{er} Octobre.

A la fin de l'année, il y a mille vaccinés et quatre mille non vaccinés.

Pendant l'année on a enregistré cent dix atteintes morbides chez les vaccinés et huit cents quatre vingt dix chez les non vaccinés.

Les rapports sont : $110/1.000$ soit 11 % chez les vaccinés, et $890/4.000$ soit 22,3 % chez les non vaccinés.

Ce résultat paraît favorable à la vaccination : 11 % contre 22,3 %.

Mais tous les sujets n'ont pas été, pendant des périodes égales, exposés à la contamination :

A partir d'Avril : $5.000 - 300 = 4.700$ restent exposés .

A partir de Juillet : $4.700 - 600 = 4.100$ restent exposés.

A partir d'Octobre : $4.100 - 100 = 4.000$.

Il y a, en cours d'année, passage du groupe des non vaccinés dans le groupe des vaccinés.

Le calcul aurait dû être fait comme suit :

1° Les 5.000 sujets sont non vaccinés pendant le premier trimestre de l'année soit le quart de l'année, ce qui fait :

$(5.000 \times 1/4)$ d'année d'exposition à la contamination de ce groupe de 5.000.

2° Pendant le deuxième trimestre 4.700 sujets appartiennent à ce groupe, c'est-à-dire : $5.000 - 300$ qui n'ont été vaccinés que le 1^{er} Avril, ce qui fait :

$(4.700 \times 1/4)$ d'année d'exposition à la contamination de ce groupe de 4.700.

3° Pendant le troisième trimestre, 4.100 personnes appartiennent à ce groupe, c'est-à-dire : $4.700 - 600$, qui n'ont été vaccinées que le 1^{er} Juillet, ce qui fait :

$(4.100 \times 1/4)$ d'année d'exposition à ce groupe de 4.100.

4° Enfin, pendant le quatrième trimestre, 4.000 per-

sonnes appartiennent à ce groupe, c'est-à-dire : 4.100 — 100, qui n'ont été vaccinées que le 1^{er} Octobre, ce qui fait :

$(4.000 \times 1/4)$ d'année d'exposition à la contamination de ce groupe de 4.000.

L' « Année-personne », (« Person-year ») d'exposition à la contamination sera de :

$(5.000 \times 1/4) + (4.700 \times 1/4) + (4.100 \times 1/4) + (4.000 \times 1/4) = 4.450$, et le rapport des atteintes morbides de: $890/4.450$. Soit 20 %, c'est-à-dire l'équivalent de 20 atteintes morbides sur 100 personnes et par an.

De même l' « année-personne » d'exposition dans les groupes vaccinés est de :

$(0 \times 1/4) + (300 \times 1/4) + (900 \times 1/4) + (1.000 \times 1/4) = 550$, car il n'y a pas de vaccinés au 1^{er} trimestre, il y a 300 vaccinés au deuxième trimestre, il y a 900 vaccinés au troisième trimestre, il y a 1.000 vaccinés au quatrième trimestre.

Le pourcentage des atteintes morbides chez les vaccinés sera donc de :

$$110/550 = 20 \%$$

Les vaccinés et les non vaccinés ont le même pourcentage d'atteinte, donc, dans ce cas, le vaccin ne met pas à l'abri de la contamination, comme les chiffres de 11 % et de 22,3 % nous le laissaient entendre .

Nous n'avons comparé que des éléments en tous points comparables, « Like with Like ». Chaque lot des individus mis en observation a été exposé au risque pendant la même période. Il reste encore, pour nous rapprocher de la vérité à tenir compte du résultat que nous aurions obtenu si nos sujets en étude n'avaient en aucune façon été exposés au risque.

3. — ABSENCE DE L'EXPOSE AU RISQUE' OU STANDARD DE COMPARAISON

La discrimination doit encore tenir compte de l'absence de l'exposé au risque.

Après des enquêtes très sérieuses, il a été démontré que les accidents d'automobiles étaient imputables dans la proportion de $3/4$ contre $1/4$, à l'absorption d'alcool par le conducteur dans les quelques heures qui ont précédé l'accident.

La déduction que l'alcool contribue aux risques d'accidents n'est pas confirmée par ces chiffres.

Le rapport de trois alcooliques sur un non alcoolique ne peut être interprété que si l'on compare ce rapport au chiffre des conducteurs qui boivent de l'alcool et qui ne causent pas d'accidents. (standard de comparaison.)

Supposons par exemple qu'il y ait 1.000 conducteurs sur une route et que 48 accidents soient enregistrés.

De ces 48 conducteurs impliqués dans ces accidents les trois quarts ont consommé de l'alcool dans les heures qui ont précédé, soit : 36, et le quart n'en a pas consommé, soit : 12.

Si les trois quarts des mille conducteurs avaient consommé de l'alcool dans les heures qui ont précédé la conduite et le quart n'en avait pas consommé, le nombre des « exposés aux risques » serait de :

$$1.000 \times 3/4 = 750 \text{ et } 1.000 \times 1/4 = 250.$$

Le pourcentage d'accidents sera donc de :

$$36/750 = 48 \% \text{ et } 12/250 = \text{aussi } 48 \%$$

La connaissance de l'« exposé aux risques » ou rap-

port consommateurs d'alcool sur non consommateurs, dans une liste de tous les conducteurs est essentielle pour nous permettre d'incriminer ce rapport dans les causes d'accidents dues à l'absorption d'alcool dans les heures qui ont précédé la conduite.

Il semble, lorsque ces discriminations successives ont été effectuées, que le praticien doive établir des résultats statistiques se rapprochant beaucoup du rôle qu'on leur demande : la traduction exacte d'un état de fait.

Il faut cependant se méfier encore d'une erreur, d'autant plus fréquente que les causes qui la conditionnent sont souvent très obscures : c'est l'abus des associations directes ou indirectes de faits qui, au premier abord, semblent dépendants les uns des autres alors qu'il n'en est rien.

4. — ASSOCIATIONS DIRECTES OU INDIRECTES

En ne tenant pas compte des variations de virulence des germes, des cycles saisonniers, ou des conditions de vie différentes, la discrimination s'avère insuffisante. C'est ce que Bradford Hill appelle des associations directes ou indirectes.

« On a observé un parallélisme entre le nombre des décès par cancer et la vente des bananes en Angleterre et au pays de Galles.

Personne (aussi loin qu'il m'en souviennne), n'a déduit de cette parenté dans le temps une relation de cause à effet.

Une telle déduction ne serait pas plus illogique que celle qui aurait été inspirée par le nombre de

moderne.

Très souvent cependant, la concordance dans le temps est employée comme évidence nouvelle de cause à effet.

Par exemple :

1° Les effets du bactériophage sont évalués en comparant les cas de choléra dans deux régions, l'une dans laquelle le bactériophage a été distribué, l'autre dans laquelle le bactériophage n'a pas été distribué, et servant de « standard comparaison ».

Dans la première, la fréquence des cas est observée pendant une période donnée, et on remarque une diminution des cas de choléra par rapport à la fréquence de ces cas, dans les années précédentes, ainsi que la fréquence des cas dans la région servant de « standard comparaison ».

Il est clair qu'il existe une association à la fois dans le temps et dans l'espace, entre la fréquence du choléra et l'administration du bactériophage.

Est-ce une association de cause à effet ?

Nous pouvons répondre que les résultats obtenus sont parfaitement compatibles avec cette hypothèse, mais cette compatibilité n'est pas l'équivalent d'une preuve.

La fréquence des cas de maladies épidémiques varie pour des raisons inconnues, et les pourcentages anormalement bas des cas de choléra dans la région dans laquelle le bactériophage fut administré, peut être le résultat d'influences de causes naturelle, indéterminées, agissant en même temps que l'expérience en cours.

La répétition de cette expérience dans une autre

région, avec des résultats équivalents, renforcerait davantage l'hypothèse que le bactériophage est salubre dans la prophylaxie du choléra.

2° L'histoire de la scarlatine doit être rappelée, illustrée par le témoignage de R. J. Graves (a system of clinical medicine) Dublin 1843.

Dans les premières années du dix-neuvième siècle, la maladie « fit de grands ravages dans Dublin » et était « très grave ».

Après l'année 1804, « elle prit un type très bénin » et était « rarement accompagnée de dangers jusqu'à l'année 1831 ».

En 1834, la scarlatine prend de nouveau le type d'une « épidémie destructive ».

La basse mortalité après 1804 était « citée chaque jour comme reflétant l'un des plus triomphants exemples de l'efficacité des nouvelles méthodes de traitement ».

Mais Graves admet que « l'expérience dérivée de l'épidémie 1834-1835 a complètement réfuté ce raisonnement et a prouvé qu'en dépit de nos vains perfectionnements, nous n'avions pas été plus heureux en 1834-35 que ne l'avaient été nos prédécesseurs en 1801-1802.

(Cité de Charles Creighton's « History of epidemics in Britain, Cambridge university press 1894, vol. II, pages 722 et 725).

3° Pendant une épidémie, des sujets sont, sur leur propre demande, vaccinés, et supportent une plus basse atteinte morbide que celle des non vaccinés.

Ce résultat concorde clairement avec l'hypothèse que la vaccination est salubre. Cependant l'associa-

tion peut être une association indirecte.

Supposons que les groupes vaccinés et non vaccinés soient égaux par l'âge, le sexe, et la durée du temps d'exposition au risque ; il est possible par ailleurs que ces groupes soient différenciés par d'autres caractères.

Ceux qui, volontairement, se sont soustraits à la vaccination, peuvent être des sujets qui prennent d'autres mesures de précaution pour éviter l'infection : ils évitent les théâtres et les cinémas pendant la période épidémique par exemple.

De plus ils peuvent appartenir à une classe plus élevée dans l'échelle sociale et être moins exposés aux risques d'infection du fait qu'ils sont mieux nourris et qu'ils vivent dans des conditions de moindre encombrement.

En comparant les sujets vaccinés et non vaccinés, nous devons toujours considérer de près que nous sommes en fait en train de comparer du semblable avec du semblable, mise à part la notion de vaccination. »

4° Un autre exemple nous est fourni par tous les modes de traitement préconisés dans le paludisme.

Pour les uns, l'administration de la quinine doit se faire trois fois par semaine.

Pour d'autres, deux jours de suite.

Pour d'autres, à doses massives.

Pour d'autres, à doses très fractionnées.

Les différents auteurs, à la suite d'expériences thérapeutiques, ont pensé que certains modes d'administration quinquine agissaient avec plus d'efficacité

sur les hématozoaires.

Exemple : En Janvier, Février Mars 1917, à Salonique, on observe des accès palustres à vivax ou præcox chez des sujets. On administre à ces malades des traitements quinquiques particuliers et variés et on note l'influence de ces traitements sur les accès. Suivant les résultats de certains modes d'administration de la quinine on change de dose ou de moment.

A la suite d'un mode particulier d'administration de la quinine, on remarque une disparition de l'accès fébrile et on en déduit une supériorité de ce traitement sur ceux qui avaient été utilisés jusqu'alors.

En réalité, la disparition des accès était due à une éclipse saisonnière du paludisme.

(M.G. Senevet. Thèse Médecine, Alger 1918.)

En criminologie, on a cherché à savoir quelles étaient les tares psychiques ou physiques qui prédisposaient le plus au crime.

On a invoqué tour à tour toutes les defficiencies psychiques, en particulier le fait que la plupart des criminels sont des illettrés. On a cherché à étayer sur des statistiques de ce genre, toutes sortes d'interprétations théoriques du crime. Spencer critique ce genre de déductions et nous met en garde contre les associations abusives que l'on a tendance à faire entre le crime et tout ce qu'un criminel peut présenter de caractères.

« Du nombre des illettrés certaines personnes ont conclu que l'ignorance est cause du crime. Il ne leur vient pas à l'idée de se demander si d'autres statistiques ne les conduiraient pas à dire que le crime est causé par l'absence d'ablutions et de linge propre,

ou par la mauvaise ventilation des logements ou le défaut de chambres à coucher séparées. Bien peu de criminels ont de quoi se changer ; la plupart logent au fond des cours ; presque tous les criminels des villes sortent des bouges ; de même ils sont souvent alcooliques. »

Ce n'est pas tout. La discrimination doit encore aider le praticien dans les prélèvements, choisis ou pris au hasard, qui seront à la base de sa statistique. Pour cela il faut préciser ce qu'on entend par : La notion de choix des échantillons.

5. — LA NOTION DE CHOIX DES ECHANTILLONS

Il existe deux sortes de statistiques :

- Les statistiques à portée générale ;
- Les statistiques à portée restreinte.

Dans les premières, il faut proscrire de façon formelle le choix des échantillons. C'est là un élément fondamental de la technique des statistiques. Dans une statistique à portée générale, il ne faut pas relever les chiffres qui concernent les agglomérations de cas sélectionnés : hôpitaux, maisons de santé, sanatoria, dispensaires.... etc...

Pour expérimenter un traitement, il est nécessaire de l'administrer à toutes les catégories d'individus et non pas seulement à des hommes ou à des femmes, à des vieillards ou à des enfants. Le meilleur procédé qui élimine complètement le facteur personnel est de ne retenir, pour l'établissement de la statistique, que un sur deux ou un sur trois des sujets qu'on nous présente. C'est le hasard seul qui doit présider au choix des éléments nécessaires à l'établissement d'une statistique.

Les statistiques à portée restreinte, à condition que les conclusions qu'on en tire soient limitées, s'accroissent d'un choix volontaire, réfléchi. C'est le cas des statistiques qui tendent à démontrer les avantages du pneumothorax dans la tuberculose unilatérale, ou le pourcentage des guérisons spontanées des tuberculoses ostéo-articulaires chez les indigènes syphilitiques.

Dans ces deux cas nos déductions doivent rester limitées et nos conclusions ne doivent absolument pas être étendues à la tuberculose pulmonaire en gé-

néral, ou à la tuberculose ostéo-articulaire sans distinction de terrain.

Nous avons remarqué, dans la littérature médicale française, le nombre considérable de statistiques qui sont faites à partir des élèves des écoles. Il est certain qu'au cours d'une vaccination jennérienne obligatoire, il est facile de faire des cuti-réactions ou de mesurer des rates. Mais cette commodité ne doit pas nous faire oublier que les enfants des écoles sont des sujets d'une catégorie qui n'est pas absolument représentative de la totalité des enfants de telle ou telle région. Il en est de même pour les statistiques que les médecins militaires publient chaque année, après les visites médicales d'incorporation. Là non plus des hommes de 21 ans ne représentent pas l'ensemble de la population et les résultats que donnent ces statistiques doivent rester limités à la seule évaluation sanitaire d'hommes adultes et jeunes.

Nous avons rappelé dans les chapitres précédents, les erreurs les plus communes qui ressortissaient à des causes volontaires et à un défaut de discrimination.

D'autres erreurs sont commises par l'application illogique des données fondamentales de l'arithmétique.

CHAPITRE III

Les Erreurs Mathématiques

Nous avons divisé ce chapitre en deux parties.

- 1° Le Calcul des pourcentages et des index totaux ;
- 2° De l'utilité de la connaissance de l'écart moyen pour le calcul de la moyenne la plus représentative de plusieurs nombres différents.

1°. — LE CALCUL DES POURCENTAGES ET DES INDEX TOTAUX.

Quand on a calculé des pourcentages à partir d'un rapport donné, par le dénombrement des cas recherchés et celui de la totalité des sujets observés, on veut poursuivre plus loin son travail, lui donner une portée plus grande, en calculant l'index total, c'est-à-dire le pourcentage qui stigmatiserait la fréquence de telle ou telle affection dans toute une région, dans toute une nation, voire même dans le monde entier.

Pour ce faire, certains additionnent les différents pourcentages trouvés, et placent cette somme en numérateur. Ils donnent au dénominateur le nombre qui correspond à celui des pourcentages additionnés au numérateur.

Le quotient de cette division est l'indice total pour le nombre d'individus considérés; il peut être ramené à cent individus, pour faciliter les comparaisons et devient alors un pourcentage.

A première vue, cette façon de procéder paraît correcte. Il n'en est rien.

Pourquoi ?

Tout d'abord : on arrive parfois en calculant les chances qu'un individu a de mourir de telle ou telle cause, à des chiffres tels que 160 %, ce qui est évidemment beaucoup.

La moyenne des pourcentages est chose encore plus courante, et cependant, aucun des auteurs qui emploient cette méthode dans la recherche des indices totaux ne commettraient la même erreur dans le calcul du taux moyen de deux sommes placées à des taux différents.

Exemple : 1.000 francs placés à 10 % et 100.000 francs placés à 4 % ne font pas 101.000 francs placés à 7 % mais à 4,3 %.

Tandis que : 50.000 francs placés à 10 % et 50.000 francs placés à 4 % font : 100.000 francs placés à 7 %.

On voit aisément que l'addition au numérateur de deux pourcentages différents ne se justifie que lorsque ces deux pourcentages sont en rapport avec des valeurs identiques (50.000 et 50.000).

Pourquoi ne pas agir de même dans les calculs d'ordre médical ?

Exemple 1 :

On trouve 49 réactions positives sur 220 sujets soumis à la cuti-réaction à la tuberculine. Soit un pourcentage de 22,2 %. D'autre part, on trouve 32 réactions positives sur 56 sujets également soumis à la cuti-réaction à la tuberculine, soit un pourcentage de 57,1 %.

L'index total calculé par la méthode des moyennes sera, en ramenant à 100 : $22,2 + 57,1$ sur 2 = 39,7 %, ce qui est complètement erroné.

Mais si on avait trouvé :

Exemple 2 :

49 réactions positives sur 220 sujets, soit un pourcentage de 22,2 % et 32 réactions positives également sur 220 sujets, soit un pourcentage de 14,5 %, l'index total calculé comme suit :

$22,2 + 14,5$ divisé par 2 = 18 % serait exact.

En effet, $49 + 32$ réactions positives divisées par 440 donnerait en ramenant à 100 sujets, 18 %.

L'index total, abstraction faite de l'addition d'éléments souvent dissemblables (enfants et adultes, hommes et femmes, habitants des villes et de la campagne) doit être obtenu en dénombrant au numérateur tous les cas de telle ou telle affection dont nous voulons connaître l'index, et en divisant ce numérateur par le nombre total des sujets observés.

Ainsi l'erreur que l'on commet forcément du fait que l'on ramène à 100 individus les quotients des rapports sujets atteints sur sujets observés, ne se trouve réalisée qu'une seule fois ; tandis que, si nous additionnons les causes d'erreur inhérentes à chaque pourcentage établi, le résultat de cette addition donnera évidemment un chiffre beaucoup plus éloigné de la vérité.

Par contre, quand les erreurs nécessaires (pour ramener à 100 individus le quotient d'un rapport) sont identiques pour plusieurs rapports, c'est-à-dire quand les rapports ont le même dénominateur (nombre de sujets examinés), le calcul de la moyenne de ce pourcentage est le même que celui du pourcentage en partant du rapport de tous les sujets atteints sur tous les sujets observés.

Des exemples assez nombreux pris dans les publications médicales illustreront mieux que nous ne pouvons le faire, la légitimité de nos objections.

L'INFECTION TUBERCULEUSE ETUDIEE PAR LES CUTI-REACTIONS DANS LE KSAR D'AIN-SEFRA

(versant saharien des hauts-plateaux oranais) par J. Tramini

Archives de l'Institut Pasteur de l'Algérie 1923 Tome I

Age	masculin	féminin	total	total de 1 à 5	de 6 à 15	de 1 à 15	Adultes
1 jour à 1 an	0 sur 0	0 sur 1	0 sur 1				
1 an à 2 ans	5 sur 15	3 sur 15	8 sur 30				
3 ans à 5 ans	9 sur 19	8 sur 28	17 sur 47	25/77 = 32%		96 — = 46,3%	
6 ans à 10 ans	30 sur 54	14 sur 30	44 sur 84		71/130 = 54%	207	
11 a. à 15 ans	14 sur 23	13 sur 23	27 sur 46				
Adultes de plus de 16 ans							
	42 sur 79	14 sur 21	56 sur 100				56/100 = 56%

46,3 + 56

Index tuberculinique total : — = 51 %.

Moniseur J. Tramini établit correctement dans sa publication l'index total tuberculinique des enfants âgés de 1 à 5 ans :

$$\text{Soit } 25 \times 100/77 = 32 \text{ \%}.$$

Celui de 6 à 15 ans :

$$\text{Soit } 71 \times 100/130 = 54 \text{ \%}.$$

Cet auteur reste dans l'exactitude des calculs de l'index pour établir l'index tuberculinique des enfants âgés de 1 à 15 ans.

Au lieu de faire la moyenne entre 32 % et 54 %, c'est-à-dire $(32 + 54) : 2 = 43 \text{ \%}$, l'auteur additionne les cuti-réactions des enfants de 1 à 15 ans, dénombre les sujets observés et ramène à 100 ce rapport :

$$\text{Soit : } 96 \times 100/207 = 46,3 \text{ \%}.$$

Puis il cherche le pourcentage des cuti-réactions positives chez 100 sujets adultes. Il en trouve 56 ce qui fait 56 %.

Et alors contrairement à ce que nous pouvions espérer de la suite correcte des calculs précédents, M.J. Tramini additionne 46,3 % et 56 % pour trouver par la méthode des moyennes, l'index tuberculinique total du Ksar d'Aïn-Sefra, soit :

$$\frac{46,3 + 56}{2} = 51 \text{ \%}.$$

En réalité l'index total est de :

152 cuti-réactions positives sur 307 cas observés, ce qui fait en ramenant à 100 : $15200/307 = 49 \text{ \%}$.

L'INFECTION TUBERCULEUSE
CHEZ LES INDIGENES DE MAZAGAN

par Madame E. Delanoe

Archives de l'Institut Pasteur de l'Algérie 1930

Tome VIII, page 109.

Ecole Ouvroir de fillettes musulmanes de la mission anglaise :

De 6 à 11 anssur 35 12 C.R. +

Ecole Ouvroir de fillettes musulmanes du Protectorat :

1^o classe de 9 à 13 anssur 36 13 C.R. +

2^o classe de 8 à 12 anssur 33 9 C.R. +

3^o classe de 5 à 12 anssur 29 5 C.R. +

Au total :sur 98 27 C.R. +

Index tuberculinique total des deux ouvroirs :

Sur 133.... 39 C.R. + soit : 29,3 %.

Ecole Coranique du Fkih Si Mohamed ben Mohamed (effectif : 26 garçons) :

de 5 à 13 anssur 26 7C..R +

Ecole Coranique de Si Mohamed ben Mohamed (effectif : 15 garçons) :

de 5 à 11 anssur 15 8 C.R. +

Index tuberculinique total des deux écoles .

Sur 41.... 15 C.R. + soit : 36,5 %.

Index tuberculinique total des fillettes et garçons musulmans de 5 à 13 ans et au nombre de 174 :

29,3 + 36,5

————— = 32,9 %

2

Ecole de l'alliance israélite :

Les résultats chez les fillettes de cette école se

répartissent de la manière suivante :

Elèves du cours supérieur et moyen :

de 10 à 14 anssur 28 22 C.R. +

Elèves des 3^e et 4^e classes :

de 8 à 13 anssur 67 29 C.R. +

Elèves de la petite classe :

de 4 à 10 anssur 57 21 C.R. +

3 Monitrices ayant accompagné les élèves ont reçu également les scarifications à la tuberculine :

de 15 à 17 anssur 3 3 C.R. +

Total :sur 155 75 C.R. +

Soit : 48 %.

L'index de la population enfantine indigène globale de Mazagan (musulmane et israélite) sur un total de 329 est donc de :

$$\text{I.T.T. : } \frac{48,3 + 32,9}{2} = 40,6 \%$$

LES FEMMES

Filles soumises indigènes du dispensaire municipal :
comprenant 75 musulmanes et 9 israélites.

de 16 à 30 anssur 89 71 C.R. +

Soit : 81,5 %.

....En résumé l'index tuberculinique total de l'élément indigène musulman et israélite (enfants et femmes de la ville de Mazagan) sur un effectif de 418 est de :

$$\frac{40,6 + 81,5}{2} = 61 \%$$

Dans la publication de Madame E. Delanee sur l'in-

fection tuberculeuse chez les indigènes de Mazagan, l'erreur est accrue par le nombre des moyennes effectuées entre les différents index tuberculiniques provenant déjà de moyennes réalisées à partir de pourcentages trouvés (dans les différents groupements d'individus de Mazagan), par cette méthode erronée.

— Les fillettes musulmanes des deux ouvroirs de Mazagan accusent un pourcentage de 29,3 % ;

— Les garçonnetts musulmans, un pourcentage de 36,5 %.

(1) Index total = moyenne des deux pourcentages précédents : soit 32,9 %.

— L'école de l'alliance israélite accuse un pourcentage de 48,3 %.

(2) Index total des fillettes et garçonnetts tant musulmans qu'israélites : 40,6 %, moyenne de l'index 32,9 % et de 48,3 %.

Madame E. Delanoe conclut alors à un index de 40,6 % dans la population enfantine de Mazagan.

Poursuivant plus loin ses calculs cet auteur trouve chez les filles soumises un pourcentage de 81,5 %. Il établit alors l'index total de Mazagan, c'est-à-dire celui des enfants, filles et garçonnetts, et des filles soumises !... soit 61 %.

Critiques :

C'est d'abord une erreur d'objectiver par le chiffre de 61 %, les nombres de cuti-réactions positives chez des éléments aussi peu comparables que les jeunes enfants des ouvroirs de Mazagan et les filles soumises. Mais cela a été amplement expliqué au chapitre de la discrimination insuffisante.

C'est ensuite l'erreur du calcul des moyennes :

En admettant que nous voulions obtenir quand même l'index total des filles soumises et des enfants de Mazagan, nous devons procéder de la manière suivante :

— Nombre total des cuti positives = 200

— Nombre total des sujets examinés = 418

d'où en ramenant à 100 : 47 %.

Nous voyons que ce chiffre correct de 47 % s'écarte de 14 % du chiffre de 61 % trouvé par l'auteur.

Le calcul des index un et deux cités dans la publication, sont également erronés. Ils sont en effet de 54 cuti-réactions sur 174 sujets, soit 31 % pour les enfants musulmans de Mazagan au lieu de 32,9 %, et de 129 cuti-réactions positives sur 329 sujets examinés, soit 39 % pour les enfants musulmans de Mazagan plus les enfants israélites, au lieu de 40,9 %.

Nous voyons par la comparaison de tous ces résultats, que du mode de calcul dépend la valeur des pourcentages sur lesquels doit s'appuyer toute action prophylactique qui veut être efficace.

L'INFECTION TUBERCULEUSE
CHEZ LES INDIGENES MUSULMANS
DU GOURARA

(Sahara oranais - 2^e note)
par R. Guéguen.

Archives de l'Institut Pasteur d'Algérie 1925
Tome III (pages 291 et 292).

« Nous avons poursuivi, au cours de l'hiver dernier, l'étude de la cuti-réaction à la tuberculine, commencée en 1923. Les résultats obtenus complètent ceux que nous avons publiés dans le tome II, fascicule IV de ces archives.

Nos recherches, cette année ont été localisées dans deux villages importants de l'annexe de Gourara : Timimoun d'une part et Ouled Saïd d'autre part, grosse oasis située à 18 kilomètres au nord de Timimoun dont elle est séparée par la Sebkha. Les sujets que nous avons observés sont des arabes Zénètes et des Harratins en majorité du sexe masculin.

Les résultats obtenus, sur un total de 578 individus, sont résumés dans les tableaux suivants, séparément pour Timimoun et pour Ouled Saïd.

VILLAGE DE TIMMOUN (région saharienne)

Population : arabe, berbère, négroïde.

Race	0 à 1 an	1 à 5 ans	6 à 15 ans	Adultes	
Arabes	1 sur 5	1 sur 1	20 sur 37	11 sur 19	soit 33/ 62
Zénètes	1 sur 4	10 sur 27	16 sur 30	12 sur 14	soit 39/ 75
Harratins ..	4 sur 21	7 sur 41	22 sur 53	33 sur 75	soit 66/190

Au total : de 0 à 15 ans : 82 sur 219 = 37,4 %

Adultes : 56 sur 108 = 51,8 %

Et l'index tuberculinique total :

$$\frac{37,4 + 51,8}{2} = 44,6 \%$$

VILLAGE DES OULED SAID (région saharienne)

Population négroïde

Race	0 à 1. an	1 à 5 ans	6 à 15 ans	Adultes	
Zénètes	1 sur 1	3 sur 5	4 sur 12	14 sur 23	soit 22/ 41
Harratins ..	1 sur 18	8 sur 36	14 sur 53	49 sur 103	soit 72/210

Au total : de 0 à 15 ans : 31 sur 125 = 24,8 %

Adultes : 63 sur 126 = 50 %

Et l'index tuberculinique total :

$$\frac{28,4 + 50}{2} = 37,4 \%$$

En réunissant les résultats des deux villages, nous obtenons :

$$\text{Index tuberculinique total} : = \frac{44,6 + 37,4}{2} = 41 \%$$

C'est le résultat que nous avaient laissé prévoir les chiffres trouvés en 1923-1924. Cet indice tuberculinique total de 41 % est très sensiblement le même que celui qu'ont obtenu en 1913, H. Foley et R. Meslin au cours de leur voyage dans le Coura-Touat.

On peut tirer des tableaux statistiques précédents quelques conclusions, les mêmes d'ailleurs, d'une façon général, que celles des auteurs qui se sont occupés de cette question dans les diverses régions du Sahara : l'indice augmente avec l'âge dans toutes les races et il est plus élevé chez les indigènes de race blanche que chez les négroïdes « admirablement adaptés au pays, à son climat, aux plus dures privations ».

Le calcul correct de l'index tuberculinique du village de Timimoun, d'après cette publication, et en ne visant spécialement que les erreurs dues aux moyennes des pourcentages, devra être :

138 cuti-réactions positives sur 327 sujets = 42 %, au lieu de : 37,3 % + 51,8 % sur 2 = 44,6 %.

Remarquons en passant que les enfants de un à quinze ans ne sont pas en tout points comparables aux adultes pour ce qui est de la cuti-réaction et que cependant l'auteur n'a pas hésité à les additionner aux adultes dans le calcul de son index total.

Pour le village des Ouled-Saïd, l'index total de 37,4 %, obtenu par la moyenne des deux pourcentages enfants de un à quinze ans et adultes, est le mê-

me que celui que nous trouvons en calculant le rapport qui existe entre le nombre des cuti-réactions positives et le nombre de sujets examinés.

La raison en est à la presque égalité du nombre des enfants observés (125) et de celui des adultes (126).

L'index total des villages de Timimoun et des Ouleds-Saïds calculé correctement devient :

232 Cuti-réactions positives sur 578 cas observés soit, en ramenant à 100 : 40 %, au lieu de 41 %, pourcentage tiré de la moyenne entre 44,6 % et 37,4 %.

Et encore ce nombre ne signifie-t-il pas grand chose car il aboutit à comparer et à additionner des sujets différents, enfants et adultes, dont les réactions à la tuberculine ne sont pas semblables. Cette erreur a été commentée au chapitre de la discrimination insuffisante.

2° — DE L'UTILITE DE LA CONNAISSANCE
DE L'ECART MOYEN POUR LE CALCUL
DE LA MOYENNE LA PLUS REPRESENTATIVE
DE PLUSIEURS NOMBRES DIFFERENTS.

*Méthode pour le calcul de l'écart moyen
de plusieurs échantillons différents.*

Toujours à propos des moyennes, un problème se pose pour le calcul de la valeur moyenne de plusieurs nombres différents.

L'habitude d'additionner les différents nombres que l'on a trouvés au cours d'expériences ou de mesures, et d'en rechercher la valeur moyenne, nous apparaît défectueuse quant à la valeur représentative du quotient de la division. Bradford Hill nous montre, à partir de vingt observations de tension artérielle, la différence qui existe entre le chiffre représentant la moyenne de ces tensions, et celui qui est calculé par la méthode exposée ci-dessous :

- 1° Après avoir additionné les différentes valeurs trouvées, il faut calculer le nombre qui représente la valeur moyenne arithmétique « mean ».
- 2° Calculer les différences de chaque valeur trouvée avec cette moyenne.
- 3° Elever au carré ces différences.
- 4° Additionner ces carrés.
- 5° En rechercher la valeur moyenne.
- 6° Calculer la racine carrée de cette dernière.

Cette racine carrée représente l'écart, en plus ou en moins, par lequel la moyenne arithmétique pré-

cédemment établie se rapproche des tailles les plus représentatives du lot étudié.

En d'autres termes, lorsqu'on a trouvé la moyenne des tailles, il faut lui ajouter, puis lui soustraire, la valeur de l'écart moyen, pour obtenir la marge la plus typique de l'ensemble des tailles étudiées.

1° Exemple :

Nous avons mesuré la taille de dix individus pris au hasard. Nous voulons connaître entre quelle taille maximum et minimum se trouve les tailles les plus représentatives du lot.

Tailles	Différences entre les tailles et la moyenne: 152	Carrés de ces différences
160	+ 8	64
170	+ 18	324
180	+ 28	784
150	— 2	4
140	— 12	144
120	— 32	1024
100	— 52	2704
200	+ 48	2304
110	— 42	1764
190	+ 38	1444
S. : 1.520	s. alg.: 00 non alg. 280	S. : 10.560
Moy. : 152	moy. : 28	Moy. : 1056 racine de 1056 = 32,4.

La moyenne arithmétique des dix tailles considérées est de 152.

La moyenne des différences au carré est de 1056.

La racine carrée de cette moyenne est de 32,4.

La taille maximum sera : $152 + 32,4 = 184,4$.

La taille minimum sera : $152 - 32,4 = 119,6$.

La taille la plus représentative du lot considéré sera comprise entre 1 mètre 196 et 1 mètre 844.

Si nous avons calculé la somme des différences qui existent entre les différentes tailles et leur moyenne, sans tenir compte du signe algébrique, nous aurions trouvé 280, et la moyenne de ces différences aurait été $280/10 = 28$.

La taille maximum aurait été de : $152 + 28 = 180$.

La taille minimum aurait été de : $152 - 28 = 124$.

La taille la plus représentative du lot considéré aurait été comprise entre 1 m. 240 et 1 m. 800, au lieu de 1 m. 196 et 1 m. 844.

Ces résultats illustrent les avantages de la méthode préconisée par Bradford Hill.

Ces avantages se montrent encore plus nettement, si nous supposons dans un premier temps que, sur les dix sujets précédemment étudiés, il y en ait cinq qui mesurent un mètre, et dans un deuxième temps qu'il y en ait cinq qui mesurent 1 m. 51 et cinq 1 m. 49.

Dans les deux cas la moyenne arithmétique est 1 m. 50.

Pourrait-on soutenir que les individus des deux groupes soient comparables, bien que la moyenne de leurs tailles soit la même ?

Si nous employons la méthode de l'écart moyen, nous avons :

— Ecart moyen dans le premier cas :

La somme des tailles est de : 15.

La moyenne est de : 1,50.

La somme des carrés des différences des tailles du lot avec cette moyenne est de : 2.500, dont la moyenne est 250.

La racine carrée de cette moyenne est : 50.

L'écart moyen est donc de 0 m. 50.

Ecart moyen dans le deuxième cas :

La somme des tailles est de : 15.

La moyenne est de : 1,50.

La somme des carrés des différences des tailles du lot avec cette moyenne est de : 10 dont la moyenne est 1.

La racine carrée de cette moyenne est: 1.

L'écart moyen est donc de 0 m. 01.

Nous voyons par cet exemple que si les moyennes sont identiques et égales à 1 m. 50, les écarts moyens calculés pour chacun des cas sont cinquante fois différents.

Cet écart nous renseigne non seulement sur la moyenne des tailles, mais encore sur leur dispersion.

2° Exemple :

20 observations de tension arté- rielle maximum	Différences de chaque observa- tion avec la moyenne : 128	Carré de ces différences
98	— 30	900
160	+ 32	1024
136	+ 8	64
128	0	0
130	+ 2	4
114	— 14	196
123	— 5	25
134	+ 6	36
128	0	0
107	— 21	441
123	— 5	25
125	— 3	9
129	+ 1	1
132	+ 4	16
154	+ 26	676
115	— 13	169
126	— 2	4
132	+ 4	16
136	+ 8	64
130	+ 2	4
Somme=2.560	Somme = 0	Somme=3.674

Ce tableau nous permet de calculer : 1^o la tension artérielle maximum moyenne.

$$\frac{2560}{20} = 128$$

2° L'écart moyen qui est égal à la racine carrée de
3.674

———— soit 13,55.

20

Les chiffres de tension artérielle maximum les plus représentatifs des 20 tensions artérielles considérées sont compris entre $128 + 13,55$ et $128 - 13,55$ soit 141,55 et 114,45.

Au lieu de mesurer les différences qui existent entre les observations et leur moyenne 128, nous pouvons adopter un procédé qui simplifie le calcul, mais nécessite une correction.

Ce procédé consiste à établir les différences de chaque observation avec le nombre 100. Le calcul peut être fait mentalement. La somme des carrés de ces différences par rapport à 100 est 19.354. La moyenne de ces différences au carré est de :

19.354

———— = 967,7

20

C'est alors que la correction intervient, elle est nécessaire car nous avons calculé les différences à partir de 100 au lieu de 128. Cette correction consiste à soustraire au nombre 967,7 le carré de la différence qui existe entre 100 et 128, soit 28 au carré, soit 784. On a alors $967,7 - 784 = 183,7$.

L'écart moyen sera la racine carrée de 183,7, soit 13,55, résultat déjà obtenu par la première méthode.

On pourrait éviter l'étape du calcul permettant d'obtenir les différences, en partant de 0 au lieu de 100, c'est-à-dire en élevant au carré les valeurs des observations elles-mêmes. Ce procédé est applicable quand on n'a qu'un petit nombre d'observations.

Cette méthode consiste à :

- 1° Trouver la moyenne des observations.
- 2° Elever au carré chaque observation.
- 3° Faire la somme de ces carrés et en rechercher la moyenne.
- 4° De cette moyenne des carrés, soustraire le carré de la moyenne des observations.
- 5° La racine carrée de cette dernière valeur est l'écart moyen.

Ecart moyen =

$$V \frac{\text{somme des carrés des observations}}{\text{nombre des observations}} - \left(\frac{\text{la moyenne des observations}}{\text{observations}} \right)^2$$

LE COEFFICIENT DE VARIATION

Le coefficient de variation est indispensable pour permettre la comparaison entre les résultats qui ont été obtenus avec des unités différentes ; grâce à lui nous pouvons avoir une idée comparative entre des mesures de longueur, des mesures de temps, des pressions en centimètres de mercure.

Ce coefficient s'obtient en divisant l'écart moyen par la moyenne arithmétique des observations et en ramenant le résultat de cette division à 100.

Calculé à partir de notre premier exemple, le coefficient de variation est :

$$\frac{32 \times 100}{152} = 21$$

Dans le tableau que nous reproduisons ci-dessous et que nous avons emprunté à P.L. Mc Kinlay et A.B.

Walker dans Edinb. med. J. 1935, 42, 407, nous voyons un exemple de l'utilisation de ce coefficient de variation.

Il nous permet de comparer l'âge moyen de 566 sujets, le nombre de pulsations moyen que ces 566 sujets présentent, le chiffre moyen de leur tension artérielle maxima ainsi que celui de leur minima.

Moyenne		Ecart moyen	coefficient de variation	chiffres limites
Age	23,2	4,02	17,31	18 à 40
Pulsations	77,3	12,83	16,60	46 à 129
Tension maximum	128,8	13,05	10,13	97 à 168
Tension minimum	79,7	9,39	11,78	46 à 108

Quand on connaît l'écart moyen et le coefficient de variation, il ne reste plus pour avoir une idée précise des moyennes, qu'à se représenter autour d'un point qu'on aura choisi, la dispersion des différentes tailles les plus représentatives du lot que nous avons étudié.

Cette distribution peut être symétrique ou asymétrique ; des graphiques peuvent la représenter.

En résumé :

— Les moyennes arithmétiques sont nettement insuffisantes pour nous représenter la valeur moyenne de plusieurs échantillons.

— Il faut connaître l'écart moyen, le coefficient de variation, le type de dispersion.

Conclusion

Que les médecins le veuillent ou non, la médecine est inséparable de la statistique ; elle repose entièrement sur elle, et ceux des praticiens qui la méprisent ne réf échissent pas que la pathologie elle-même n'est qu'une vaste statistique. Quand on dit que tel symptôme est constant ou présent dans telle affection, c'est qu'on a fait une statistique.

Nous nous sommes efforcés, à la suite de statisticiens tels que Bradford Hill, Ross et Stott, etc.... de montrer au praticien les erreurs les plus courantes pour lui permettre d'établir des statistiques plus précises et moins empiriques.

Ces erreurs sont, nous l'avons vu, tout d'abord volontaires, faciles à éliminer puisque dépendantes de notre volonté et de notre conscience professionnelle.

Puis ce sont les erreurs par discriminations insuffisante : il est relativement aisé d'entraîner son esprit à la recherche des caractères communs et différentiels nous permettant de rapprocher ou de placer dans des catégories différentes, les éléments de nos statistiques.

Ensuite nous éviterons les erreurs mathématiques, en appliquant logiquement les principes de calculs nécessaires à la recherche des pourcentages ou des indices totaux.

Enfin, nous noublierons pas que les moyennes

arithmétiques ne correspondent pas toujours à la réalité et qu'il est nécessaire de connaître la notion d'écart moyen, pour avoir une idée exacte de la dispersion, autour d'un type choisi, des autres éléments en étude.

Bibliographie

- BRADFORD HILL. — *The Lancet*, Janvier, Février, Mars et Avril 1937.
- CREIGHTONS'S (Charles). — *History of epidemics in Britain*. Cambridge University press 1894. Vol. II (p. 722 and 725).
- CRAPOUILLOT (LE). — Numéro du mois de Mai 1938 (Criminologie).
- DELANOE (Madame E.) — *L'infection tuberculeuse chez les indigènes de Mazagan*. Archives de l'Institut Pasteur de l'Algérie 1930, Tome VIII (p. 109).
- DUHAMEL (Georges). — *Les Maîtres*, de la Chronique des Pasquiers.
- GUEGUEN (R.) — *L'infection tuberculeuse chez les indigènes musulmans de Gourara* (Sahara Oranais 2^e note). Archives de l'Institut Pasteur de l'Algérie, Tome III 1925 (p. 291 et 292).
- LUMIERE (A.) et VIGNES (P.) — *Avenir Médical*, XXX^e Année, N^o 2, Février 1933.
- NEUMANN. — *Traité des maladies parasitaires* (p. 691).
- ROSS ET STOTT. — *Tables of statistical errors*. The annals of tropical medicine and parasitology. T 5 24 Octobre 1911 (p. 347).
- SINCLAIR LEWIS. — *Arrowsmith*, roman.
- SENEVET (G.) — *Contribution à l'étude de la pathogénie des accès palustres*. Thèse Médecine, Alger 1918.
- TRAMINI (J.) — *L'infection tuberculeuse étudiée par les cuti-réactions à la tuberculine dans le ksar d'Aïn-Sefra (versant saharien des hauts plateaux oranais)*. Archives de l'Institut Pasteur de l'Algérie, Tome I. 1923.

LABORATOIRE DE PARASITOLOGIE

DE LA

FACULTE DE MEDECINE D'ALGER

THESE N^o 22

Vu :

Le Président de Thèse :

SENEVET

Vu le Doyen :

LIEBLANC

Vu et permis d'imprimer :

Alger, le 12 mai 1938

LE RECTEUR,

PIERRE MARTINO.

Tables des Matières

	PAGES
Avant propos	11
Historique	15
Définition	15
Méthodes de la statistique	16
Présentation des statistiques	17
Plan	17

CHAPITRE I

Les causes d'erreur volontaires	19
A — Causes d'erreur par l'application du secret profes- sionnel	20
B — Causes d'erreur par publications prématurées	23

CHAPITRE II

Causes d'erreur par discrimination insuffisante	28
1. Mélange de chiffres non comparables.....	31
2. Négligence de la période d'exposition au risque	33
3. Absence de l'exposé au risque ou standard de comparaison	37
4. Associations directes et indirectes	38
5. La notion de choix des échantillons	44

CHAPITRE III

Les erreurs mathématiques	46
1° — Le calcul des pourcentages et des index totaux..	47
2° — De l'utilité de la connaissance de l'écart moyen pour le calcul de la moyenne la plus repré- sentative de plusieurs nombres différents	62
Le coefficient de variation	68
Conclusion	71
Bibliographie	73

